JSLHR

Research Article

Efficacy of a Supplemental Phonemic Awareness Curriculum to Instruct Preschoolers With Delays in Early Literacy Development

Howard Goldstein,[a] Arnold Olszewski,[a] Christa Haring,[a] Charles R. Greenwood,[b] Luke McCune,[b] Judith Carta,[b] Jane Atwater,[b] Gabriela Guerrero,[b] Naomi Schneider,[c] Tanya McCarthy,[c] and Elizabeth S. Kelley[d]

Purpose: Children who do not develop early literacy skills, especially phonological awareness (PA) and alphabet knowledge, prior to kindergarten are at risk for reading difficulties. We investigated a supplemental curriculum with children demonstrating delays in these skills.
Method: A cluster randomized design with 104 preschool-age children in 39 classrooms was used to determine the efficacy of a supplemental PA curriculum, *PAth to Literacy*. The curriculum consists of 36 daily scripted 10-min lessons with interactive games designed to teach PA and alphabet skills. A vocabulary intervention (*Story Friends*), which also uses a small-group format, served as the comparison condition.
Results: Multilevel modeling indicated that children in the experimental condition demonstrated significantly greater gains on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) First Sound Fluency (Dynamic Measurement Group, 2006) and Word Parts Fluency (Kaminski & Powell-Smith, 2011) measures. Educational relevance was evident: 82% of the children in the experimental condition met the kindergarten benchmark for First Sound Fluency compared with 34% of the children in the comparison condition. Teachers reported overall satisfaction with the lessons.
Conclusions: Results indicated that the vast majority of children demonstrating early literacy delays in preschool may benefit from a supplemental PA curriculum that has the potential to prevent reading difficulties as children transition to kindergarten.

The National Early Literacy Panel (NELP, 2008) reports that developmental trajectories for reading skills begin early. Children who lag behind their same-age peers early in the development of literacy skills often struggle in school (Storch & Whitehurst, 2002). For example, Foster and Miller (2007) found that students who fell behind their peers in kindergarten in early literacy tasks struggled with text comprehension in third grade. If we are to improve reading skills nationally, we must develop prevention and early intervention strategies that ensure children are entering school with the skills needed to become successful readers.

Whitehurst and Lonigan (1998) proposed inside-out (code focused) and outside-in (meaning focused) skills as the two critical domains of emergent literacy. Outside-in skills refer to oral language ability as evidenced by development in contextual knowledge and semantic skills. Inside-out skills refer to understanding the phoneme and grapheme units of language. Phonological awareness (PA), particularly phonemic awareness, is a necessary precursor to fluent decoding and conventional reading (Anthony, Williams, McDonald, & Francis, 2007; NELP, 2008; Whitehurst & Lonigan, 1998). Alphabet knowledge, knowing the names and sounds of letters, and grapheme–phoneme correspondence are requisite decoding skills (NELP, 2008; Whitehurst & Lonigan, 1998). Together, alphabet and PA skills may account for more than half the variance in first-grade decoding (Lonigan, Burgess, & Anthony, 2000).

PA refers to "the ability to detect, manipulate, or analyze the auditory aspects of spoken language (including the ability to distinguish or segment words, syllables, or phonemes), independent of meaning" (NELP, 2008, p. 3). This metalinguistic skill does not seem to develop naturally (Wagner & Torgesen, 1987) and must be taught explicitly

[a]University of South Florida, Tampa, FL
[b]University of Kansas, Kansas City
[c]The Ohio State University, Columbus
[d]University of Missouri, Columbia
Correspondence to Howard Goldstein: hgoldstein@usf.edu

Editor: Sean Redmond
Associate Editor: Nicole Terry
Received December 29, 2015
Revision received March 24, 2016
Accepted May 10, 2016
DOI: 10.1044/2016_JSLHR-L-15-0451

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

(Ehri et al., 2001). Although a variety of skills such as blending, segmenting, elision, rhyming, and initial sound identification are associated with PA, it is perhaps best viewed as a single metalinguistic construct (Anthony & Francis, 2005).

Alphabet knowledge refers to the ability to name printed letters and to identify the sounds associated with them. This may be the single best predictor of later reading ability (Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Alphabet knowledge and PA are correlated, and development of one may influence development of the other (Johnston, Anderson, & Holligan, 1996). These skills are relatively stable across the preschool and early school years (Lonigan et al., 2000). Interventions that target PA and alphabet knowledge together seem to be more effective than interventions that use a whole-word approach to reading (Fielding-Barnsley, 1997). Lonigan, Purpura, Wilson, Walker, and Clancy-Menchetti (2013) found that preschoolers in a code-focused intervention made gains in PA and alphabet knowledge, whereas preschoolers receiving meaning-focused early literacy interventions did not. Effective interventions that target PA and alphabet knowledge must be made available for early childhood educators to prepare children for reading success. Commonly used preschool curricula generally are not sufficient for teaching early literacy skills to children at risk for disabilities (Goldstein, 2011), and there is a paucity of supplemental, evidence-based curricula suitable for struggling learners in early childhood (Greenwood et al., 2011). Thus, curricula are needed that are effective for teaching children demonstrating delays in early literacy development and feasible for implementation in early childhood settings using multitiered systems of supports (MTSS) to meet children's needs.

MTSS is an increasingly popular model of providing appropriate supports to children with a variety of skill levels (Berkeley, Bender, Gregg Peaster, & Saunders, 2009). Multiple tiers typically are depicted in a triangle or a pyramid. The base of the pyramid represents Tier 1, which entails a high-quality, whole-class curriculum with regular screening and assessment to identify children who are not making adequate progress. Tiers 2 and 3 are levels of support provided to children who are lagging behind in academic or behavioral skills. In the academic sphere, tiers of instruction may vary in terms of the amount of instruction, the targets of instruction, and the teaching strategies used (e.g., level of prompting, reinforcement). Tier 2 instruction typically is delivered in small groups, and Tier 3 typically is delivered one on one. Movement among tiers is informed by frequent progress monitoring. Although MTSS has only recently been adopted in early education settings, there are indications that this model of instruction is appropriate and efficacious for young children (Buysse et al., 2013; Gettinger & Stoiber, 2007; Greenwood et al., 2012; VanDerHeyden, Snyder, Broussard, & Ramsdell, 2008; VanDerHeyden, Witt, & Gilbertson, 2007).

Our goal was to find, adapt, or develop a supplemental curriculum that would fulfill several criteria (Goldstein & Olszewski, 2015). First, it should follow a developmentally appropriate scope and sequence. Second, it should be suitable for small groups of children. Third, it should be easily integrated into preschool classroom routines (e.g., center rotations) by classroom teachers or aides. Fourth, it should provide instruction appropriate for children who fit the profile of a Tier 2 candidate—that is, it should target children who are beginning to show delays in foundational reading skills compared with their peers, thus placing them at risk for developing later reading disabilities. Fifth, it should be efficacious.

Several code-focused early literacy interventions have been developed for use as supplemental instruction for struggling learners. These interventions have demonstrated efficacy for teaching skills such as alphabet knowledge, PA, print concepts, and name writing to children who have been identified as likely benefiting from supplemental instruction (Justice, Chow, Capellini, Flanigan, & Colton, 2003; Koutsoftas, Harmon, & Gray, 2009; O'Connor, Jenkins, Leicester, & Slocum, 1993; van Kleeck, Gillam, & McFadden, 1998). However, none of these interventions teach PA from larger sound units (e.g., compound words) to smaller sound units (e.g., phonemes), similar to the way PA skills are thought to develop in young children (Anthony & Francis, 2005). Also, in these intervention studies, research staff provided instruction with treatment doses that are not feasible for most preschools. Several of these studies focused specifically on children with disabilities or documented speech-language disorders (e.g., O'Connor et al., 1993; van Kleeck et al., 1998). *PAth to Literacy* is a center-based, small-group, scripted intervention that targets PA skills and alphabet knowledge, including letter names and sounds (Kruse, Spencer, Olszewski, & Goldstein, 2015). The curriculum is designed to be delivered to groups of two to three children for about 10 min/day. In an early efficacy trial of *PAth to Literacy* (Kruse et al., 2015), research staff delivered the intervention to children in Head Start classrooms. Progress was monitored using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) First Sound Fluency (FSF; Dynamic Measurement Group, 2006) and Word Parts Fluency (WPF; Kaminski & Powell-Smith, 2011) measures. Effects were evident using a multiple baseline design across small groups: Five of the seven children who completed the intervention demonstrated gains on the WPF measure, and all seven children demonstrated gains on the FSF measure. At the end of the study, all seven children scored above the kindergarten benchmark score of 10 on the FSF measure.

Despite the impressive improvements in PA, the applicability of this intervention can be questioned because members of the research team delivered it outside the classroom. For a Tier 2 intervention to be deemed viable, effects need to be demonstrated when school personnel deliver it in the classroom. Furthermore, teacher feedback (i.e., social validity) regarding feasibility is required to determine how readily the intervention may be incorporated into classrooms outside of research studies and whether implementation will be sustained (Goldstein & Olszewski, 2015). Previous research has indicated that teachers are capable of implementing language and early literacy

curricula with a high degree of procedural fidelity, although this does not often result in high-quality teaching (Justice, Mashburn, Hamre, & Pianta, 2008). That is, teachers are able to complete the tasks associated with instructional curricula but lack the flexibility to provide enhanced learning opportunities to individual children (Justice et al., 2008). The scripted nature of *PAth to Literacy*, including predetermined student feedback, may remedy problems with inconsistencies in the quality of implementation of the curriculum.

The purpose of this study was to evaluate the efficacy of a supplemental PA intervention when delivered by teachers within pre-K classrooms to children not responding to Tier 1 instruction. We sought to test the hypothesis that the *PAth to Literacy* curriculum would promote significantly larger growth in children's PA skills compared with a second group using an automated storybook language intervention (*Story Friends*) focused on promoting vocabulary and comprehension skills (Kelley, Goldstein, Spencer, & Sherman, 2015). Although there is evidence that vocabulary growth and emergence of PA are related (e.g., lexical restructuring model; Metsala & Walley, 1998), the relatively brief period of intervention in the design plan was not expected to significantly affect PA skills of children receiving the *Story Friends* intervention, thus making the comparison scientifically interesting. In addition, teachers were asked to complete a social validity survey to determine the feasibility and perceived utility of *PAth to Literacy*. The specific research questions were as *follows:*

1. Are superior FSF and WPF PA skills outcomes produced by the *PAth to Literacy* group versus the *Story Friends* group?
2. Are observed effects for the *PAth to Literacy* group moderated by pretest early literacy skills (Test of Preschool Early Literacy [TOPEL]; Lonigan, Wagner, Torgesen, & Rashotte, 2007) and language skills (Comprehension Evaluation of Language Fundamentals Preschool–Second Edition [CELF]; Wiig, Secord, & Semel, 2004) or the number of intervention sessions?
3. Do the two groups differ at posttest on a researcher-developed measure of alphabet knowledge (Letter Sound ID) and a standardized measure of PA, print, and alphabet knowledge (TOPEL)?
4. Do classroom teachers perceive the intervention as beneficial to children and feasible to implement in the classroom?

Method

Experimental Design

A cluster randomized design was used to compare the effects of *PAth to Literacy* and the *Story Friends* intervention on children's growth in PA skills. A cluster consisted of one classroom with two to three low-performing children per classroom. A total of 561 children initially participated in a multigated screening process. Of those, 423 children were excluded during the screening process on the basis of testing criteria, and 25 children were excluded for other reasons, including behavior issues, children leaving the classroom, and more than three qualifying children in a classroom. Classrooms were excluded if fewer than three children qualified. This produced 39 clusters with 113 enrolled children in all (see Figure 1). This sample exceeded the 32 clusters of three children estimated by our power analysis, allowing ample room for attrition.

Following pretesting, randomization occurred at the classroom level within sites to control for site effects. Twenty classrooms and 60 children participated in the *PAth to Literacy* intervention, and 19 classrooms and 53 children participated in the *Story Friends* comparison intervention. In some classrooms, teachers and teacher aides took turns implementing the intervention with children. The overall design of the study was organized into three phases. The first phase, from weeks 1 to 9, was a multiple-gating screening and enrollment phase. The second phase, from weeks 10 to 25, consisted of intervention exposure. The third phase, from weeks 26 to 28, evaluated the maintenance of skills following completion of instruction.

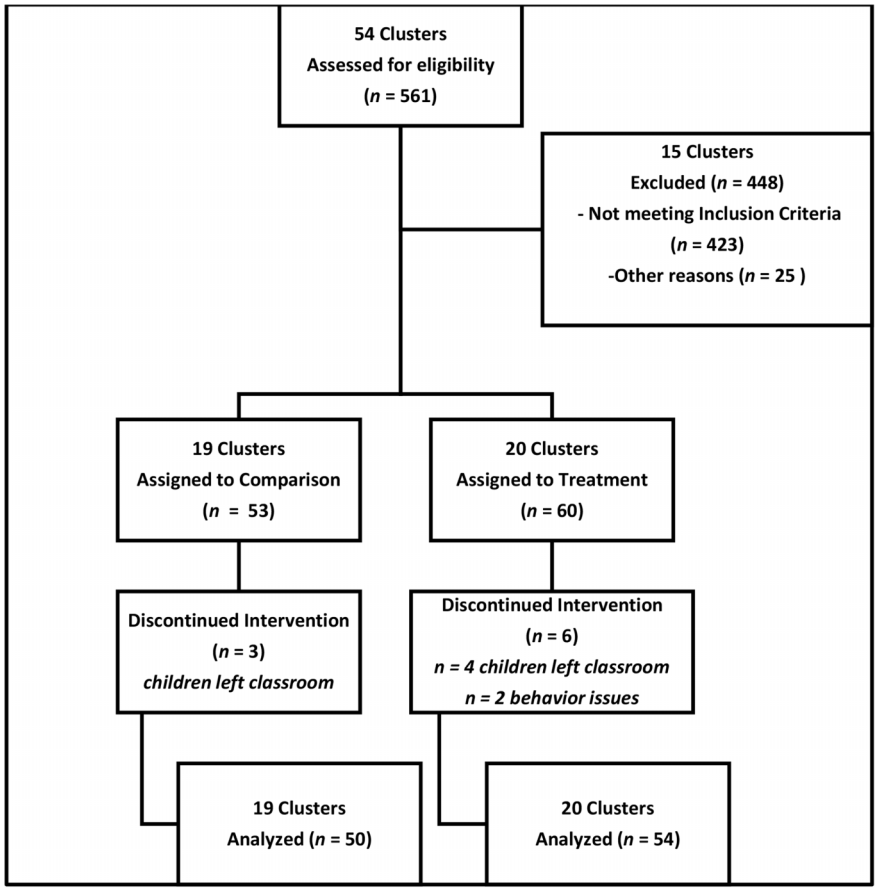
Participants

This study was conducted at three sites: Ohio, Kansas, and Florida. At each site, urban classrooms serving large proportions of minority families, often in high-poverty communities, were recruited. In Ohio and Kansas, the study was conducted in public pre-K classrooms. In Florida, the study was conducted in child care centers that served as voluntary pre-K providers. There were eight classrooms in Ohio, 11 classrooms in Kansas, and 20 classrooms in Florida. Parents were asked to identify the racial/ethnic category that best described their child. The majority of children in our sample were identified as either Hispanic (35%) or African American (33%). The remaining children were White (16%), mixed/other (13%), or Asian (3%). Five families did not identify a race/ethnicity. Parents were asked to complete surveys that asked about family size and family income. Of the 113 participating families, 89 returned the survey. Of the families who completed the survey, 47% fell below the federal poverty line for their family size.

All children with parental consent completed screening assessments, as is typical in an MTSS or Response to Intervention approach (Fuchs & Fuchs, 2006). The screening sought to identify three children who were not developing PA skills because these children were likely to benefit from additional instructional support. The selected participants exhibited basic expressive and receptive English language proficiency but deficits in PA after a period of time in the classroom and exposure to instruction.

A multiple-gating screening procedure took place between September and December. In step 1 of screening, children who scored more than 4 points on the DIBELS FSF measure or above 12 on the First Sounds IGDI

Figure 1. CONSORT table of enrollment.



(Individual Growth and Development Indicator) Fall Screening measure were excluded. The cut-point for the FSF measure was chosen to include children who may have correctly guessed the first sound of one or two items. The First Sounds IGDI cut-point was determined by the developers. In step 2 of screening, children who scored more than 4 points on the DIBELS FSF measure or below 3 on the Picture Naming IGDI measure were excluded. The Picture Naming IGDI was included to determine whether children had sufficient English proficiency to participate in instruction. In step 3, children who scored more than 99 on the TOPEL Phonological Awareness subtest were excluded because we strove to include children who performed below the mean. In classrooms in which more than three children remained in the study following the three gates of screening, the children whose test results indicated the greatest need for Tier 2 support were included.

Groups were equivalent in demographics (see Table 1) with one exception: There were significantly more boys than girls in the *PAth to Literacy* group than in the *Story Friends* group, $\chi^2(1) = 5.70$, $p < .05$. Groups were not different in mean age or on any pretest scores (i.e., FSF, WPF, First Sounds IGDI, Sound ID IGDI, Letter Sound ID, TOPEL scales, and CELF).

Setting and Procedures

Teachers or teacher aides conducted the two interventions in small groups in their classrooms during the intervention phase. Most classrooms began intervention in

Table 1. Demographic characteristics of participants in each group.

Variable	*PAth to Literacy*	*Story Friends*
Teacher/classroom–children clusters (*n*)	20	19
Children at start of intervention (*n*)	60	53
Child gender (%)		
Male	58*	36
Female	42*	64
Mean age of children at pretest (months)	56.4	55.9
Children with individualized education plans (*n*)	2	3
English language learners (*n*)	18	15
Mean CELF Core Language Index	82.2	83.4
Families below poverty line (%)	33	61

Note. One cluster = one teacher/classroom and two to three low-performing children. CELF = Clinical Evaluation of Language Fundamentals Preschool–Second Edition.

*$p = .05$.

January. Intervention sessions for both conditions lasted about 10 min each. The *PAth to Literacy* group received the intervention three to five times per week depending on classroom schedules. Children received a total of 19 to 36 lessons depending on attendance and how quickly the children in the cluster acquired skills. The *Story Friends* group received the intervention three times per week for 13 weeks. The mean number of sessions was 29 for the *PAth to Literacy* group and 35 for the *Story Friends* group. Research staff trained teachers and teacher aides, observed intervention sessions, and supported implementation.

Teacher–children clusters were randomly assigned to one of the two conditions conducted in the intervention phase. In the *PAth to Literacy* group, teacher–child clusters participated in scripted lessons targeting PA skills from that curriculum (Kruse et al., 2015). Lessons included visual materials and often incorporated gestures. Children were given frequent opportunities to respond throughout the lessons and were given scripted feedback contingent on the response of the group. The end of each lesson included a brief review, during which teachers collected data on student responses. Lessons were divided into 12 units each containing three parallel lessons. If children mastered the skills after two lessons, the cluster progressed to the next unit; otherwise, the third lesson was administered.

The *Story Friends* group also participated in small groups at listening centers using the *Story Friends: Jungle Friends* curriculum (Kelley et al., 2015). Children in small groups listened to interactive prerecorded stories that included instruction on low-frequency vocabulary words and basic concept words. Teachers in this condition were responsible for helping the children attend to the stories and encouraging responses during the automated questions. Children in the *Story Friends* group participated in three listens of a book each week. The 13-book curriculum includes an introductory book and three units that include three instructional books and one review book. Each instructional book introduces two low-frequency vocabulary words, two basic concepts words, and model comprehension questions.

Outcome Measures

A variety of PA and language measures were administered during the study. The progression is shown in Table 2. Prior to the intervention, children participated in three gates of screening and one additional gate of pretesting. About halfway through the intervention phase, progress monitoring assessments were completed. Immediately following completion of the intervention, all children were assessed using posttest measures. Maintenance assessments were conducted two to three weeks following posttesting.

FSF

The FSF measure served as the primary proximal measure of phonemic awareness (DIBELS; Dynamic Measurement Group, 2006). Slight modifications were made to the administration so that the first sound was modeled at the end of each sample item—for example, "The first sound you hear in the word *moon* is /m/." Children were asked to identify the initial phoneme in as many orally presented words as possible in a 1-min fluency measure. Children received 2 points for correctly producing the initial phoneme of a word and 1 point for producing the initial blend of a word. There are 30 items and a possible maximum score of 60. Parallel forms of the measure were used. Alternate form reliability for FSF is 0.82, and predictive validity with DIBELS Phoneme Segmentation Fluency and Nonsense Word Fluency is 0.46 to 0.51 and 0.41, respectively (Cummings, Kaminski, Good, & O'Neil, 2010).

WPF

A modified version of the DIBELS WPF measure (under development at Dynamic Measurement Group; Kaminski & Powell-Smith, 2011) served as a secondary measure of PA. Similar to the FSF measure, the instructions were modified slightly so that the first part was modeled at the end of each sample item—for example, "The first part of *sailboat* is *sail*." Children were asked to produce the first part of as many orally presented words as possible in a 1-min fluency measure. Children received 1 point every time they correctly produced the initial phoneme, initial phoneme blend, or initial syllable of the two-syllable target words. In previous studies a ceiling effect was noted for this measure. Therefore, in the present study, multiple forms were combined such that the maximum score was 36 rather than 18. Reliability and validity data are not available because this measure is under development.

First Sounds IGDI

The First Sounds IGDI 2.0 (McConnell, Bradfield, & Wackerle-Hollman, 2014; Wackerle-Hollman, Schmitt, Bradfield, Rodriguez, & McConnell, 2015) is a measure of PA, particularly initial phoneme awareness. The examiner presented a card depicting two to three pictures, named the pictures, and then asked the child to point to the picture that started with the target phoneme. This untimed assessment included 30 items. Children received 1 point for each correct response for a maximum score of 30. Internal consistency on the basis of congeneric reliability was reported to be 0.76, and concurrent construct validity correlation with the TOPEL Phonological Awareness subtest was reported to be 0.61 (Bradfield, McConnell, Rodriguez, & Wackerle-Hollman, 2013).

Sound ID IGDI

The Sound ID IGDI 2.0 (McConnell et al., 2014) served as a distal measure of alphabet knowledge. This is a 15-item measure in which the examiner presents a phoneme and asks children to choose the one that matches the phoneme from a field of three letters on a card. This measure is untimed, and children get 1 point for each correct response for a maximum of 15 points. Internal consistency on the basis of congeneric reliability was 0.81, and concurrent construct validity correlation with the TOPEL Phonological Awareness subtest was 0.71 (Bradfield et al., 2013).

Table 2. Measures used throughout the study by phase and testing week.

Measure	Screening			Intervention			
	Screen 1 (week 1)	Screen 2 (week 5)	Screen 3 (week 7)	Pretest week 9)	Progress (week 19)	Posttest (week 25)	Maintenance (week 28)
FSF	X	X		X	X	X	X
WPF		X		X	X	X	X
First Sounds IGDI	X	X		X	X	X	X
Rhyme IGDI	X	X		X	X	X	X
Sound ID IGDI				X		X	
Letter Name ID				X		X	X
Letter Sound ID				X		X	X
TOPEL PA			X			X	
TOPEL PK			X			X	
CELF				X		X	

Note. FSF = DIBELS First Sound Fluency; WPF = DIBELS Word Parts Fluency; IGDI = Individual Growth and Development Indicator; TOPEL = Test of Preschool Early Literacy; PA = Phonological Awareness subtest; PK = Print Knowledge subtest; CELF = Clinical Evaluation of Language Fundamentals Preschool–Second Edition.

TOPEL

The Phonological Awareness and Print Knowledge subtests of the TOPEL (Lonigan et al., 2007) were administered at pretest and posttest. These subtests were used as distal measures of PA and concepts of print (including alphabet knowledge). The subtests of this standardized, norm-referenced assessment have a mean of 100 and an *SD* of 15. The alpha reliability coefficients range from .87 to .96, and criterion validity estimates range from .59 to .77 (Lonigan et al., 2007).

Letter and Sound Identification Mastery Monitor

The Letter and Sound Identification Mastery Monitor is a researcher-developed measure of alphabet knowledge. This measure was used at pretest, posttest, and maintenance testing to monitor whether children learned the names and sounds of the 11 letters introduced in the *PAth to Literacy* curriculum. The examiner presented the child with a card depicting the target letter. Children were asked "What letter is this?" and "What sound does this letter make?" Children earned 1 point for each correct letter name and 1 point for each correct letter sound for a total of 22 points. This curriculum-based measure served as the proximal measure of alphabet knowledge.

CELF

The CELF provided a descriptive measure of child language. This standardized, norm-referenced assessment has a mean of 100 and an *SD* of 15. Core Language Index scores were calculated from scores on the Sentence Structure, Word Structure, and Expressive Vocabulary subtests. This assessment was administered at pretest. The internal consistency ranges from .73 to .96, and test–retest reliability ranges from .77 to .92 (Wiig et al., 2004).

Implementation Fidelity

Training of *PAth to Literacy* teachers was conducted in small-group sessions lasting approximately 3 hr. During these sessions, members of the research team demonstrated the intervention, showed sample video clips, distributed training manuals and intervention materials, and helped teachers practice delivering lessons. Teachers kept training manuals and videos to practice independently. Several weeks after the training, members of the research team met individually with teachers and performed a standard checkout procedure to ensure that teachers were ready to begin implementing the intervention with children. Additional support and training were provided to teachers who struggled during the checkout process. Upon completion of checkout, teachers began implementing the *PAth to Literacy* intervention in their classrooms. Each teacher was observed and coached by a member of the research staff at least once during their first 3 days of implementation. Coaching consisted of researchers and individual teachers discussing areas in which fidelity of implementation was low. Because teachers had little difficulty delivering the intervention with high fidelity, no systematic method for coaching was utilized.

Teachers in the comparison condition participated in small-group training sessions lasting approximately 2 hr. Due to the automated nature of the *Story Friends* intervention, the training was shorter. *Story Friends* teachers also received weekly observations and support from a member of the research staff.

The research team conducted weekly observations of the intervention to assess fidelity of implementation. A researcher-developed observation checklist contained eight items that were scored using frequency criteria: (a) preparing children for lessons, (b) reading scripted lessons verbatim, (c) using visual materials, (d) correctly saying words and sounds, (e) providing correct feedback, (f) fluent progress through lessons, (g) accurate data recording, and (h) keeping children's attention. Teachers could earn a possible 18 points for appropriately implementing all items. The average number of observations for classrooms in the *PAth to Literacy* condition was nine. Prior to the start of the study, the research staff completed training on the observation

checklist and practiced scoring fidelity of implementation from videotaped lessons from a prior pilot study. To complete training, each researcher scored at least two videos with 90% interrater reliability with the second author. If agreement was below 90%, training continued and the sessions were rescored until interrater reliability was above 90% on two separate videos.

Overall, the fidelity of implementation was high (84%). Fidelity scores ranged from 46% to 100%. Lower scores typically corresponded with observations that occurred at the beginning of the study, immediately after winter break, or lessons in which the instructional language was different from previous lessons (e.g., lesson 7 introduced first sound identification). Teachers responded favorably with minimal coaching from the research staff.

Fidelity of Assessment and Scoring Reliability

Research staff completed rigorous checkout procedures for each measure prior to administration. To examine fidelity, assessments were audio-recorded, and trained research assistants rated a random sample (at least 20% from each wave of assessment) using a fidelity checklist specific to each measure. Mean fidelity scores for each measure were 95% or higher.

A trained member of the research team scored all measures. For the DIBELS and Letter Sound Mastery Monitor measures, at least 20% of assessments were blindly rescored by a separate trained member of the research team for purposes of evaluating scoring reliability. An item-by-item comparison was used to determine agreement percentages; the total number of agreements was divided by the total number of agreements plus disagreements and multiplied by 100. Interobserver agreement means for FSF and WPF were 96% (range = 25%–100%) and 98% (range = 75%–100%), respectively. For FSF, the 25% agreement was an isolated incident in which the child responded only four times and was difficult to understand, thus resulting in a scoring discrepancy for three of the four responses. TOPEL, CELF, and IGDI measures were not assessed for scoring reliability due to the nature of the measures; these measures involved picture pointing tasks that were not possible to capture via audio recording.

Social Validity

Upon completion of the intervention, the *PAth to Literacy* teachers were asked to complete a 22-item Likert-type survey regarding their satisfaction with the intervention and training materials. Surveys were collected from the teacher primarily involved in administering the intervention in each classroom. Teachers responded on a scale of 1 (*strongly disagree*) to 6 (*strongly agree*) to positive statements regarding the intervention. Questions were grouped into categories: (a) adequacy of training, (b) perceived child benefits, (c) ease of lesson delivery, (d) overall feasibility of the curriculum in the classroom, and (e) likelihood to make modifications.

Statistical Analysis

Analyses included 39 classroom clusters and 104 children. Nine children were dropped due to attrition ($n = 7$) and behavior issues ($n = 2$; see Figure 1). To address the research questions, multilevel growth models were calculated separately for FSF, WPF, and First Sounds IGDI scores. First, the pattern of growth was assessed on each variable to determine whether linear or quadratic growth would be more appropriate. Second, the differences in the mean intercept and slope by groups were evaluated. Third, moderation of groups' growth by CELF pretest scores, TOPEL Phonological Awareness and Print Knowledge pretest scores, attendance, child gender, English language learner status, and individualized education plan status was evaluated.

Results

The observed means and standard deviations for both groups are shown in Table 3. Because FSF and WPF were both positively skewed, count-based variables with a variance substantially larger than the mean, negative binomial multilevel regression was applied instead of traditional regression on the basis of assumptions of normality (Agresti, 2007). This change of distribution on these dependent variables allowed for a more accurate modeling through generalized linear mixed modeling but with a different interpretation of the parameters themselves. These FSF and WPF estimates in growth models represent the natural log increase in the count of the dependent variable for each unit increase in the appropriate independent variable. This can be changed to a multiplicative or percentage increase in the dependent variable for each unit increase in the appropriate independent variable by taking the exponent of the estimate.

DIBELS FSF

A linear trend of the natural log of the counts was determined to be the most appropriate model for growth in FSF because the quadratic term did not significantly contribute information to the model; likelihood ratio (LR) ($df = 7$) < .001, $p = .999$. Intraclass correlation coefficients (ICCs) in the analyses for children and classrooms were .054 and .001, respectively, indicating that differences between children explained 5.4% of the variance in FSF posttest scores, whereas differences in classrooms explained less than 1% of the variance in posttest scores. There was a significant effect of group on growth such that children in the *PAth to Literacy* group grew 26.6% faster on average than children in the *Story Friends* group ($\beta = 0.244$, $SE = 0.096$, $p = .011$; see Table 4). Children in the *PAth to Literacy* group also demonstrated 3.33 times higher predicted FSF scores at maintenance than children in the *Story Friends* group ($\beta = 1.203$, $SE = 0.193$, $p < .001$). This corresponds to a small effect of *PAth to Literacy* on growth and intercept according to Cohen's (1988) f^2 effect size on the basis of relative increase in pseudo multiple correlation squared. Attendance was not a significant moderator of

Table 3. Descriptive statistics by phase, group, and time.

Measure	Week	*PAth to Literacy* (*n* = 54)		*Story Friends* (*n* = 50)		Effect size (Cohen's *d*)
		M	*SD*	*M*	*SD*	
FSF	1	0.06	0.23	0.18	0.63	
	5	0.37	0.96	0.24	0.75	
	9	0.57	1.19	0.55	1.60	
	19	3.08	6.12	1.92	4.15	
	25	14.78	9.50	5.06	7.40	1.14
	28	15.52	11.23	7.43	8.60	0.80
WPF	5	1.33	3.33	1.00	2.75	
	9	2.81	4.37	3.69	5.77	
	19	8.57	7.76	5.37	6.69	
	25	11.74	7.11	7.98	7.30	0.52
	28	14.33	7.17	11.65	8.93	0.33
First Sounds IGDI	5	17.35	4.06	17.31	3.04	
	9	16.78	4.83	17.68	3.90	
	19	18.31	4.11	18.35	3.57	
	25	20.39	4.86	20.13	5.18	0.05
	28	21.83	5.37	20.77	4.54	0.21
Letter Sound ID	Pre	3.56	3.54	2.80	2.99	
	Post	8.13	3.07	6.08	3.99	0.58
Sound ID IGDI	Pre	8.90	3.20	7.65	3.27	
	Post	8.76	3.66	7.94	3.35	0.23
TOPEL PA	Pre	84.57	7.51	83.96	9.85	
	Post	94.24	12.03	94.69	10.74	–0.04
TOPEL PK	Pre	93.13	12.86	93.82	11.38	
	Post	97.80	13.02	97.45	15.67	0.02
CELF Core Language Scale score	Pre	81.78	13.30	83.14	12.83	
	Post	86.20	11.82	88.49	11.30	–0.20
Total sessions		28.81	3.59	34.78	3.38	

Note. FSF = DIBELS First Sound Fluency; WPF = DIBELS Word Parts Fluency; IGDI = Individual Growth and Development Indicator; TOPEL = Test of Preschool Early Literacy; PA = Phonological Awareness subtest; PK = Print Knowledge subtest; CELF = Clinical Evaluation of Language Fundamentals Preschool–Second Edition.

Table 4. Results for multilevel growth models using DIBELS measures.

Variable	DIBELS First Sound Fluency				DIBELS Word Parts Fluency			
	Estimate	*SE*	*p*	f^2	Estimate	*SE*	*p*	f^2
Intercept	**1.965**	**0.141**	**<.001**	—	–2.595	1.686	.124	—
Wave[a]	**0.922**	**0.070**	**<.001**	**.575**	**0.831**	**0.045**	**<.001**	**.130**
Group[b]	**1.203**	**0.193**	**<.001**	**.022**	**4.870**	**1.585**	**.002**	**.012**
Wave × Group[c]	**0.244**	**0.096**	**.011**	**.011**	—	—	—	—
Attendance	–0.014	0.023	.553	.002	0.051	0.036	.162	.001
Attendance × Condition[c]	—	—	—	—	**–0.136**	**0.049**	**.005**	**.024**
English language learner	0.149	0.184	.418	.002	0.247	0.188	.188	.077
Individualized education plan	–0.835	0.432	.053	.008	–0.146	0.435	.737	.001
Female	**–0.489**	**0.172**	**.004**	**.037**	0.128	0.172	.458	.054
CELF pre score	**0.018**	**0.008**	**.020**	**.009**	**0.015**	**0.008**	**.044**	**.027**
TOPEL PA pre score	0.000	0.011	.985	.000	0.019	0.011	.079	.015
TOPEL PK pre score	**0.016**	**0.007**	**.020**	**.020**	0.000	0.007	.963	.061

Note. Effect sizes (f^2) of 0.02 or below correspond to small effects, effect sizes around 0.15 correspond to medium effects, and effect sizes of 0.35 or higher correspond to large effects (Cohen, 1988). Bold rows indicate effects that are statistically significant, $p < .05$.CELF = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; TOPEL = Test of Preschool Early Literacy; PA = Phonological Awareness subtest; PK = Print Knowledge subtest.

[a]Wave variable is centered at wave 7 such that intercept represents the end of the study. [b]The *PAth to Literacy* experimental condition is compared with the *Story Friends* comparison condition. [c]Interaction is included in the model only where it is significant ($p < .05$).

any effects, and inclusion of covariates did not substantially change the growth trajectory or effect of condition on growth (see Table 4). In practical terms, however, 82% of the children in the *PAth to Literacy* group at maintenance met or exceeded the beginning of kindergarten benchmark for FSF (10) compared with only 34% of the children in the *Story Friends* group, with mean scores of 15.5 versus 7.4. The effect sizes that were based solely on the posttest and maintenance test scores were $d = 0.99$ and 0.75, respectively.

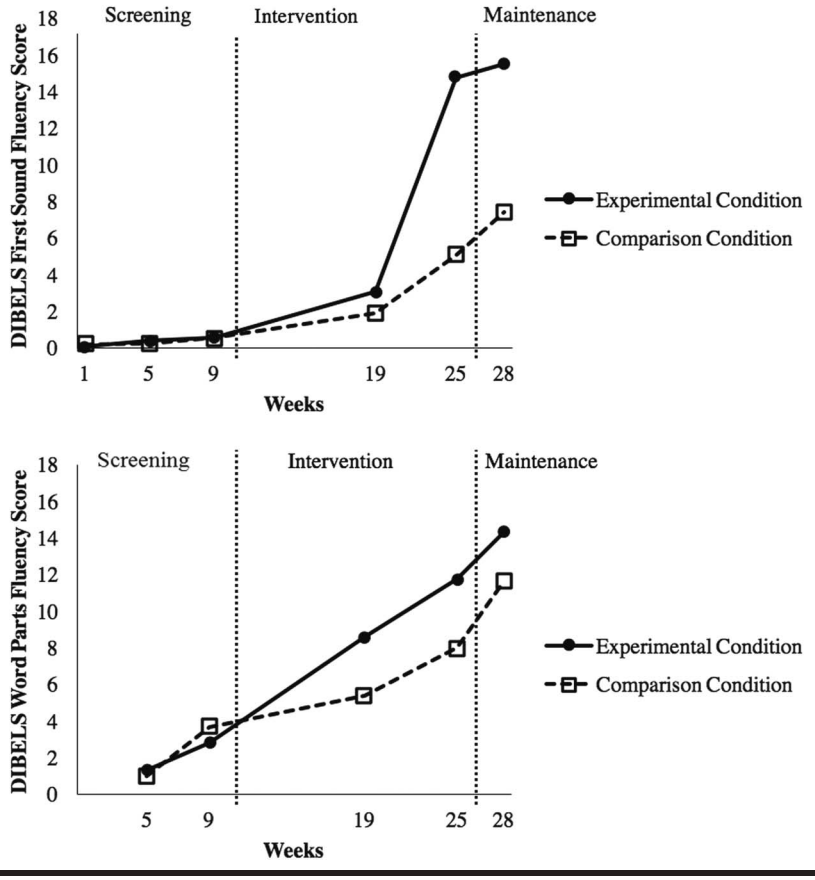
DIBELS WPF

A linear pattern of growth for the log counts of WPF also was used (see Figure 2). ICCs in the analyses for children and classrooms were .152 and .132, respectively, indicating that differences between children explained 15.2% of the variance in WPF posttest scores and differences in classrooms explained 13.2% of the variance in posttest scores. The groups did not differ significantly in the relative rate of change in WPF ($\beta = -0.019$, $SE = 0.092$, $p = .832$), although children are predicted to have 82% higher WPF scores in the *PAth to Literacy* group versus the *Story Friends* comparison group on average across time. Including attendance as a moderator indicated a significant interaction of attendance with the effect of *PAth to Literacy* on the intercept ($\beta = -0.136$, $SE = 0.049$, $p = .005$) such that more attendance at respective sessions decreased the difference between the conditions in WPF scores by 12.7% for each additional session. Up until a child has attended the average number of sessions (i.e., 32), the *PAth to Literacy* condition still significantly produced higher WPF scores; there is no significant difference between the two conditions for children who attended more than the average number of sessions. Including the set of covariates did not lead to substantial changes in the parameters (see Table 4).

First Sounds IGDI

First Sounds IGDI scores were approximately normally distributed, and thus normal-theory multilevel growth modeling was applied. The addition of a fixed quadratic trend (i.e., one that is assumed to be equal across children) significantly improved the model's fit to the data; LR ($df = 1$) = 22.369, $p < .001$. However, there was no need to make this a random trend (i.e., one that is allowed to vary across children) because the model would not be significantly improved by such added complexity; LR ($df = 6$) = 0.4779, $p < .998$. ICCs in the analyses for children and classrooms were .014 and .403, respectively, indicating that differences between children explained 1.4% of the variance in First

Figure 2. Condition means on First Sound Fluency and Word Parts Fluency across time.



Sounds IGDI posttest scores, whereas differences in classrooms explained 40.3% of the variance in posttest scores.

After accounting for the covariates and thereby the variance that would otherwise be considered error in the model, there was a significant interaction of condition with linear growth ($\beta = 1.320$, $SE = 0.535$, $p = .014$) and quadratic growth ($\beta = 0.199$, $SE = 0.098$, $p = .043$) such that children in the *PAth to Literacy* group grew faster and with more acceleration than children in the *Story Friends* group (see Table 5). Although statistically significant, both effect sizes were small, indicating that these relationships did not represent an educationally important difference given that the predicted scores differed by only a maximum of less than 2 points. Attendance did not contribute significantly to the model.

Letter and Sound Identification Mastery Monitor

In the two-level (i.e., children nested within classrooms) pre- and posttest regression model predicting posttest Letter Sound ID scores, there was no significant main effect of groups after accounting for pretest scores. Also, there were no other significant interaction or covariate effects. The classroom ICC was .365, indicating that 36.5% of the variance in Letter and Sound ID was due to classroom characteristics.

IGDI Sound ID

Using the two-level regression model with IGDI Sound ID as the dependent variable, there was no significant main effect for groups, and addition of covariates reduced the significance of the interaction between pretest and groups. Attendance was not a significant moderator of the interaction or a main effect predicting IGDI Sound ID. The classroom ICC was .279, indicating that 27.9% of the variance in Sound ID was due to classroom characteristics.

Table 5. Results for multilevel growth models using First Sounds IGDI.

Variable	First Sounds IGDI			
	Estimate	*SE*	*p*	f^2
Intercept	5.569	4.436	.213	—
Wave[a] (linear)	**1.325**	**0.388**	**.001**	**.009**
Wave[a] (quadratic)	0.124	0.070	.080	.023
Group[b]	1.323	1.068	.218	.001
Wave × Group[c]	**1.320**	**0.535**	**.014**	**.006**
Wave[a] × Group[c]	**0.199**	**0.098**	**.043**	**.003**
Attendance	0.024	0.084	.777	.005
English language learner	1.237	0.654	.062	.006
Individualized education plan	–1.180	1.457	.420	.006
Female	–1.141	0.616	.067	.032
CELF pre score	0.030	0.027	.270	.008
TOPEL PA pre score	0.058	0.038	.132	.014
TOPEL PK pre score	**0.080**	**0.025**	**.002**	**.053**

Note. Effect sizes (f^2) of 0.02 or below correspond to small effects, effect sizes around 0.15 correspond to medium effects, and effect sizes of 0.35 or higher correspond to large effects (Cohen, 1988). Bold rows indicate effects that are statistically significant, $p < .05$. CELF = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; TOPEL = Test of Preschool Early Literacy; PA = Phonological Awareness subtest; PK = Print Knowledge subtest.

[a]Wave variable is centered at wave 7 such that intercept represents the end of the study. [b]The *PAth to Literacy* experimental condition is compared with the *Story Friends* comparison condition. [c]Interaction is included in the model only where it is significant ($p < .05$).

TOPEL

Tests of effects introduced in a stepwise manner revealed that there is no significant main groups effect or interaction effects on the TOPEL Phonological Awareness posttest after accounting for pretest. Inclusion of attendance and the covariates in the model did not change this relationship. The classroom ICC was .386, indicating that 38.6% of the variance in TOPEL Phonological Awareness subtest scores was due to classroom characteristics. Likewise, for TOPEL Print Knowledge, there were no significant main or interaction effects. The classroom ICC was .360, indicating that 36.0% of the variance in TOPEL Print Knowledge scores was due to classroom characteristics.

Social Validity

Overall, *PAth to Literacy* teachers' satisfaction was high. From high to low, the mean category ratings on a 6-point scale were (a) adequacy of training ($M = 5.1$, $SD = 0.9$), (b) perceived child benefits ($M = 5.0$, $SD = 1.2$), (c) ease of lesson delivery ($M = 4.9$, $SD = 1.2$), (d) overall feasibility of the curriculum in the classroom ($M = 4.8$, $SD = 1.4$), and (e) likelihood to make modifications ($M = 4.2$, $SD = 1.6$). Individual items with the lowest ratings were "The PA lesson activities were engaging to my students" ($M = 4.4$), "The PA lessons could be easily included in my class schedule at least three times per week" ($M = 4.5$), and likely to make modifications to the curriculum ($M = 4.2$). Teachers in Kansas tended to be less satisfied with the curriculum than teachers in Ohio and Florida. In particular, they were dissatisfied with the amount of time required to implement lessons each day and noted that the children seemed bored and frustrated with the lessons. This indicates a need to make the lessons more engaging.

Discussion

The purpose of the study was to evaluate effects of a supplementary Tier 2 intervention to determine its suitability for application within an MTSS approach to preschool services. Children who were found to be not responsive to core classroom instruction through systematic universal screening and progress monitoring were enrolled in the study. Classroom clusters serving these children were randomized to receive two alternative interventions. One targeted PA skills (*PAth to Literacy*) and the other targeted vocabulary and comprehension skills (*Story Friends*).

With regard to the first two research questions, children in the *PAth to Literacy* group demonstrated accelerated growth on the DIBELS FSF measure compared with children

in the *Story Friends* group. The Cohen's *d* effect size postintervention was large. Likewise, children in the *PAth to Literacy* group also demonstrated higher scores on the DIBELS WPF at posttest than children in the *Story Friend* group, although the difference was not significant in a generalized linear mixed model. This indicates that children acquired phonemic awareness skills best when taught via *PAth to Literacy*. Tests of moderator effects did not show expected effects of TOPEL and CELF pretest scores or the number of intervention sessions. For example, neither attendance nor TOPEL or CELF pretest scores moderated the primary effect on the FSF measure.

With regard to the third research question, group differences on the Letter Sound ID Mastery Monitor and TOPEL Phonological Awareness and Print Knowledge subtests were not significant at posttest. Children in both conditions demonstrated gains on these skills from pretest to posttest, although group differences in growth rates were not significant. This indicates that *PAth to Literacy* may not boost the learning of alphabet skills beyond classroom instruction. Furthermore, PA skills did not generalize to a broad standardized measure.

With regard to the fourth research question, results of the social validity measure were encouraging. Educators gave the highest ratings to the adequacy of training, perceived child benefits, ease of lesson delivery, and overall feasibility of the curriculum in the classroom. The ratings showed some inclination to make modifications to the curriculum to fit classroom routines and individual child needs. A suggestion was to make lessons more like games to keep the children from getting bored.

The most promising finding was the effect *PAth to Literacy* had on FSF scores. This measure is a general outcome measure with a history of good sensitivity to growth in PA development that has shown good reliability for pre-K and kindergarten students (Cummings et al., 2010). General outcome measurement is based on identifying a single task that provides an indication of change in the general outcome desired. General outcome measures are brief, easy to collect, and psychometrically sound indices that describe current levels of achievement and rates of progress (Fuchs & Deno, 1991). The practical significance of this finding was evident in that the vast majority of children in the experimental condition (82%) met or exceeded the beginning of kindergarten benchmark for FSF compared with 34% of the children in the comparison condition. These results were all the more impressive because repeated testing and a multigating procedure that monitored progress of pre-K children from September through December were used to identify children with delays in early literacy development. The participants clearly demonstrated a lack of progress in learning PA skills from the general classwide curriculum. In fact, FSF continued to average less than 1 for both conditions, with growth occurring only after first sound identification was introduced in the *PAth to Literacy* curriculum.

The effect of *PAth to Literacy* on FSF is not large until posttest (see Figure 2). This is likely due to the sequence of instruction throughout the curriculum. The first half of the curriculum focuses on earlier developing PA skills (i.e., blending and segmenting at the syllable level). This instruction aligns better with the WPF measure, thus explaining gains in WPF at week 19. The second half of the curriculum introduces initial sound identification, thus explaining the effect on FSF at posttest (week 25).

Other criterion measures showed less impressive results. Differences in WPF are evident with moderate effective sizes at posttest and maintenance ($d = 0.51$ and 0.33, respectively). However, these effect sizes must be interpreted with caution because as the Wave × Condition interaction was not significant in the generalized linear mixed model. WPF is an earlier developing skill that often shows improvement in the early stages of intervention and evidently from the general curriculum as well. Although the multilevel growth model revealed a significant condition difference for the First Sounds IGDI, the small effect size and the magnitude of difference in conditions' means indicated that a clinically substantial difference was absent. Because this measure requires children to select from two pictures, it seems that the large chance component yields a relatively insensitive measure of PA growth.

Little experimental effect was shown for measures of alphabet knowledge skills. Both conditions showed improvements, and posttest results showed medium ($d = 0.56$) but nonsignificant effects on letter–sound identification in the multilevel model. It appears that the intervention shows limited effects beyond the effects of other educational experiences. More research on how to ensure mastery of letter–sound correspondence is warranted. The literature seems to have little evidence of robust effects in this area. For example, Piasta and Wagner (2010) conducted a meta-analysis of 27 multicomponent alphabet intervention studies (e.g., alphabet plus PA) and calculated overall average weighted effect sizes of .43 for letter name knowledge and .65 for letter sound knowledge. The modest effect sizes may be due to the reliance on rote memorization for alphabet knowledge, the decreased focus on alphabet in multicomponent studies, and the fact that children in comparison conditions tend to be exposed to the alphabet at home and in the classroom (Piasta & Wagner, 2010).

Gains were demonstrated on the TOPEL Phonological Awareness and Print Knowledge subtests by both conditions; the mean improvements on the Phonological Awareness subtest were about 1 *SD* for both conditions. The mean improvements on the Print Knowledge subtest were less than 0.5 *SD*. Improvements may be attributed to the general curriculum and the increased emphasis on PA and print awareness. For example, classrooms participating in the Florida voluntary pre-K program require that children demonstrate proficiency on a school readiness assessment to maintain funding. Such accountability efforts related to early literacy skills may help explain why children in the comparison condition demonstrated impressive gains on the TOPEL. The lack of a condition effect for the TOPEL scores also may be due to poor alignment between the lessons and this measure. For example, the two primary PA skills

targeted on the TOPEL are blending and elision. Although blending is taught in *PAth to Literacy*, it does not target phoneme-level blending required for many items on the TOPEL. Furthermore, elision is not directly taught in *PAth to Literacy*. Generalized short- and long-term effects need to be explored on PA and reading measures.

The results of this study are particularly impressive because the primary outcome measures (FSF and WPF) required children to produce the initial part or phoneme of a word without cues. In contrast, previous intervention studies relied primarily on measures that required matching or picture pointing (Byrne & Fielding-Barnsley, 1991; Justice et al., 2003). Comparison to studies that utilized initial phoneme production tasks is difficult because these studies utilized single-case design (Koutsoftas et al., 2009) or included children with speech-language disorders (van Kleeck et al., 1998).

Although the majority of children responded to treatment, six children scored below 5 points (our inclusion criteria) on the FSF measure at posttest and maintenance. This indicates that approximately 11% of children were nonresponders. Moderator variables and anecdotal accounts were not sufficient for identifying specific factors that accounted for the lack of growth in these children. This percentage is lower than many other studies of early literacy (Al Otaiba & Fuchs, 2002). On the basis of a review of 23 studies of preschoolers to third graders, 8% to 80% of children did not respond to treatment depending on the measure; the most common deficits were in PA. In a full-scale MTSS model, the 11% of children in the present study would be ideal candidates for Tier 3 intervention. Future research should identify factors that may help identify children who will not respond to Tier 2 interventions.

The present study is unique in that it framed the intervention within an MTSS framework for preschoolers specifically demonstrating delays in early literacy skills. This study represents a strong test of a supplementary curricula approach to teaching a developmental progression of PA skills: blending, segmenting, word part identification, and first sound identification. A multigating procedure monitored progress of pre-K children from September through December to identify children who clearly demonstrated a lack of progress in learning PA skills from the general classwide curriculum. The PA intervention was implemented during a typical pre-K activity for 10 to 15 min per day for an average of 29 sessions.

The comparison condition represented a similar small-group instructional format focusing on vocabulary teaching, which Metsala and her colleagues have hypothesized to benefit the development of PA skills as part of their lexical restructuring theory (Metsala & Walley, 1998; Walley, Metsala, & Garlack, 2003). The fact that 34% of the children in the comparison condition met the benchmark for the beginning of kindergarten may indicate a generalized benefit of vocabulary instruction but more likely is the result of the general curriculum.

Although modest gains were shown for alphabet knowledge and distal measures of PA, the robust effect on DIBELS FSF is notable. This is significant because this PA measure clearly requires children to respond at the phonemic level. Another notable feature of this study was that teachers and paraeducators administered the scripted intervention in the normal course of preschool activities. These educators were able to manage small groups of children and provide contingent feedback on the basis of the groups' performance on PA and letter–sound tasks. Overall, the fidelity of implementation was high (84%).

Limitations and Future Research

The first limitation of this study is the lack of alignment between instruction and assessment. The scope of the PA instruction was larger than the measures used. For example, four distinct PA skills were introduced in *PAth to Literacy*: blending, segmenting, initial syllable identification, and initial phoneme identification. However, only two of these skills were targeted via the FSF and WPF outcome measures. The TOPEL targets blending (some items at the phoneme level, which was not included in *PAth to Literacy*) and elision (not taught in *PAth to Literacy*). Although these skills form the larger construct of PA (Anthony & Lonigan, 2004), assessment of distinct tasks would provide insight into the development of PA and the efficacy of instruction. Earlier effects of instruction may have been observed in the study had measures of blending and segmenting been used.

The second limitation of the study is that the choice of the comparison condition was of a different intervention not thought to affect PA. Although comparing two interventions allows a more rigorous evaluation of the experimental condition, there is a chance that children in the comparison condition made gains due to extra attention, exposure to oral language skills, or repeated testing. Nevertheless, the extended screening period used in this study lends support to the notion that identified Tier 2 children did not seem to be making progress through business-as-usual instruction.

A third limitation of the study is that resources were not sufficient to measure the alternative outcomes of the two conditions. For example, vocabulary growth was not monitored as closely for the comparison condition as in previous studies (Goldstein et al., 2016). Nevertheless, a brief posttest vocabulary mastery monitor was administered to children across conditions to determine how many of the words introduced in *Story Friends* children were able to define. On average, children in the comparison condition were able to define 13.8 of the 18 words taught via *Story Friends*. Children in the experimental condition defined an average of only 2 of the 18 words at posttest. These findings indicate that the children in the comparison condition benefited from vocabulary instruction.

In addition to addressing these limitations, future research should investigate the implementation of this intervention within a full-scale MTSS model. The goal of this study was to investigate the efficacy of the specific intervention as delivered by teachers. Nevertheless, teachers

were not responsible for assessment and decision making, as would be the case in a full-scale MTSS model. Although we do not suspect that the minimal amount of teacher coaching provided by researchers had a significant effect on child outcomes, there is a need to investigate how well teachers implement the intervention without researcher support. Furthermore, because high fidelity of implementation does not always equate to high-quality instruction, instructional quality may improve if teachers are allowed to adapt the intervention to suit the needs of their classroom. It is hypothesized that aligning the intervention with Tier 1 classroom instruction and Tier 3 supports for treatment nonresponders will result in improved child learning.

Future research should investigate whether children maintain the PA skills acquired during intervention. Furthermore, there is a need to investigate whether these skills generalize to improved reading outcomes during the school years. For students identified as treatment nonresponders, additional research may help pinpoint specific variables that affect children's response to early literacy intervention.

Educational Implications

Overall, this study demonstrated the efficacy of a supplementary PA intervention for teaching initial phoneme identification—an important preliteracy skill. The fact that all but 18% of children in the experimental condition met the kindergarten benchmark indicates educational significance of the intervention. Children are not expected to meet this benchmark until the following school year. This suggests that children who require Tier 2 supports may catch up to their peers following a brief but intensive small-group intervention. The intervention was judged by teachers to be feasible and useful in the classroom. Thus, this intervention may soon be used in educational settings in efforts to prevent children from developing reading disabilities.

Another important implication of this study is the use of a multiple-gating screening procedure to identify candidates for supplementary instruction. Many previous studies ignore the identification process and instead focus on larger populations that may be at risk. The multiple-gating procedure, in which children's progress is monitored over the course of a semester through brief language and literacy screening measures, seems to efficiently identify children who are truly at risk for literacy problems. The measures used in this study are available to educators, and a similar process may help educators monitor children in their classrooms and provide appropriate supplementary interventions to support struggling children.

Early childhood education can present a number of challenges to effective instruction. Potential challenges include high turnover in personnel, child care providers with limited education, varying philosophies on pedagogy and the importance of an academic focus, and inconsistent Tier 1 curricular quality. It often may be unrealistic to expect teachers to provide multiple tiers of instruction. This scripted intervention has the potential to supersede many of these challenges. The fact that mainly paraeducators were able to implement training with fidelity and obtain good outcomes in about 12 weeks is an indication of the viability of *PAth to Literacy* as a Tier 2 intervention.

Acknowledgment

This work was supported by Center for Response to Intervention in Early Childhood Cooperative Agreement R324C080011 from the Institute of Education Sciences, U.S. Department of Education.

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*(5), 300–316. doi:10.1177/07419325020230050501

Agresti, A. (2007). *An introduction to categorical data analysis*. New York, NY: Wiley.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255–259. doi:10.1111/j.0963-7214.2005.00376.x

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43–55. doi:10.1037/0022-0663.96.1.43

Anthony, J. L., Williams, J. M., McDonald, R., & Francis, D. J. (2007). Phonological processing and emergent literacy in younger and older preschool children. *Annals of Dyslexia, 57,* 113–137. doi:10.1007/s11881-007-0008-8

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of Response to Intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*(1), 85–95. doi:10.1177/0022219408326214

Bradfield, T., McConnell, S. R., Rodriguez, M., & Wackerle-Hollman, A. (2013). *Summary of psychometric characteristics for second-generation individual growth and development indicators for universal screening* [Unpublished technical report]. University of Minnesota, Minneapolis.

Buysse, V., Peisner-Feinberg, E. S., Soukakou, E., LaForett, D., Fettig, A., & Schaaf, J. M. (2013). Recognition & response: A model of response to intervention to promote academic learning in early education. In V. Buysse & E. S. Peisner-Feinberg (Eds.), *Handbook for response to intervention in early childhood* (pp. 69–84). Baltimore, MD: Brookes.

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83,* 451–455. doi:10.1037/0022-0663.83.4.451

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cummings, K. D., Kaminski, R. A., Good, R. H., & O'Neil, M. (2010). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of first sound fluency. *Assessment for Effective Intervention, 36,* 94–106. doi:10.1177/1534508410392209

Dynamic Measurement Group. (2006). *First Sound Fluency–Experimental Version*. Eugene, OR: Author.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287. doi:10.1598/rrq.36.3.2

Fielding-Barnsley, R. (1997). Explicit instruction in decoding benefits children high in phonemic awareness and alphabet knowledge. *Scientific Studies of Reading, 1,* 85–98. doi:10.1207/s1532799xssr0101_5

Foster, W. A., & Miller, M. (2007). Development of the literacy achievement gap: A longitudinal study of kindergarten through third grade. *Language, Speech, and Hearing Services in Schools, 38,* 173–181. doi:10.1044/0161-1461(2007/018)

Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99. doi:10.1598/RRQ.41.1.4

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–501.

Gettinger, M., & Stoiber, K. (2007). Applying a response-to- intervention model for early literacy development in low-income children. *Topics in Early Childhood Special Education, 27*(4), 198–213. doi:10.1177/0271121407311238

Goldstein, H. (2011). Knowing what to teach provides a roadmap for early literacy intervention. *Journal of Early Intervention, 33,* 268–280. doi:10.1177/1053815111429464

Goldstein, H., Kelley, E. S., Greenwood, C. R., McCune, L., Carta, J., Atwater, J., . . . Spencer, T. (2016). Embedded instruction improves vocabulary learning during automated storybook reading among high-risk preschoolers. *Journal of Speech, Language, and Hearing Research, 59,* 484–500.

Goldstein, H., & Olszewski, A. (2015). Developing a phonological awareness curriculum: Reflections on an implementation science framework. *Journal of Speech, Language, and Hearing Research, 58*(6), S1837–S1850. doi:10.1044/2015_JSLHR-L-14-0351

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The response to intervention (RTI) approach in early childhood. *Focus on Exceptional Children, 43*(9), 1–22.

Greenwood, C. R., Carta, J. J., Atwater, J., Goldstein, H., Kaminski, R., & McConnell, S. R. (2012). Is a response to intervention (RTI) approach to preschool language and early literacy instruction needed? *Topics in Early Childhood Special Education, 33*(8), 48–64. doi:10.1177/0271121412455438

Johnston, R. S., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in pre-readers: The nature of the relationship. *Reading and Writing, 8,* 217–234. doi:10.1007/BF00420276

Justice, L. M., Chow, S. M., Capellini, C., Flanigan, K., & Colton, S. (2003). Emergent literacy intervention for vulnerable preschoolers: Relative effects of two approaches. *American Journal of Speech-Language Pathology, 12,* 320–332. doi:10.1044/1058-0360(2003/078)

Justice, L. M., Mashburn, A., Hamre, B., & Pianta, R. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23,* 51–68. doi:10.1016/j.ecresq.2007.09.004

Kaminski, R., & Powell-Smith, K. A. (2011). *Word Parts Fluency*. Eugene, OR: Dynamic Measurement Group.

Kelley, E. S., Goldstein, H., Spencer, T., & Sherman, A. (2015). Effects of automated Tier 2 storybook intervention on vocabulary and comprehension learning in preschool children with limited oral language skills. *Early Childhood Research Quarterly, 31,* 47–61. doi:10.1016/j.ecresq.2014.12.004

Koutsoftas, A. D., Harmon, M. T., & Gray, S. (2009). The effect of tier 2 intervention for phonemic awareness in a response-to-intervention model in low-income preschool classrooms. *Language, Speech, and Hearing Services in Schools, 40,* 116–130. doi:10.1044/0161-1461(2008/07-0101)

Kruse, L. G., Spencer, T. D., Olszewski, A., & Goldstein, H. (2015). Small groups, big gains: Efficacy of a tier 2 phonological awareness intervention with preschoolers with early literacy deficits. *American Journal of Speech-Language Pathology, 24,* 189–205. doi:10.1044/2015_AJSLP-14-0035

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36,* 596–613. doi:10.1037/0012-1649.36.5.596

Lonigan, C. J., Purpura, D. J., Wilson, S. B., Walker, P. M., & Clancy-Menchetti, J. (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. *Journal of Experimental Child Psychology, 114,* 111–130. doi:10.1016/j.jecp.2012.08.010

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool Early Literacy*. Austin, TX: Pro-Ed.

McConnell, S. R., Bradfield, T. A., & Wackerle-Hollman, A. K. (2014). Early childhood literacy screening. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 141–170). Washington, DC: American Psychological Association.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Retrieved from http://lincs.ed.gov/publications/pdf/NELPReport09.pdf

O'Connor, R. E., Jenkins, J. R., Leicester, N., & Slocum, T. A. (1993). Teaching phonological awareness to young children with learning disabilities. *Exceptional Children, 59,* 532–546.

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly, 45,* 8–38. doi:10.1598/RRQ.45.1.2

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265–282. doi:10.1037/0022-0663.96.2.265

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38,* 934–947. doi:10.1037/0012-1649.38.6.934

VanDerHeyden, A. M., Snyder, P. A., Broussard, C., & Ramsdell, K. (2008). Measuring response to early literacy intervention with preschoolers at risk. *Topics in Early Childhood Special Education, 27*(4), 232–249. doi:10.1177/0271121407311240

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*(2), 225–256. doi:10.1016/j.jsp.2006.11.004

van Kleeck, A., Gillam, R. B., & McFadden, T. U. (1998). A study of classroom-based phonological awareness training for preschoolers with speech and/or language disorders. *American Journal of Speech-Language Pathology, 7,* 65–76. doi:1058-0360/98/0703-0065

Wackerle-Hollman, A. K., Schmitt, B. A., Bradfield, T. A., Rodriguez, M. C., & McConnell, S. R. (2015). Redefining individual growth and development indicators: Phonological awareness. *Journal of Learning Disabilities, 48,* 495–510. doi:10.1177/0022219413510181

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212. doi:10.1037/0033-2909.101.2.192

Walley, A. C., Metsala, J. L., & Garlack, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing, 16,* 5–20.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69,* 848–872. doi:10.1111/j.1467-8624.1998.tb06247.x

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition*. San Antonio, TX: Pearson Assessments.